Generalized information-entropy measures and Fisher information

Marco Masi[*]

via Varese 16, 20121, Milan, Italy.

February 6, 2008

Abstract

We show how Fisher's information already known particular character as the fundamental information geometric object which plays the role of a metric tensor for a statistical differential manifold, can be derived in a relatively easy manner through the direct application of a generalized logarithm and exponential formalism to generalized information-entropy measures. We shall first shortly describe how the generalization of information-entropy measures naturally comes into being if this formalism is employed and recall how the relation between all the information measures is best understood when described in terms of a particular logarithmic Kolmogorov-Nagumo average. Subsequently, extending Kullback-Leibler's relative entropy to all these measures defined on a manifold of parametrized probability density functions, we obtain the metric which turns out to be the Fisher information matrix elements times a real multiplicative deformation parameter. The metrics independence from the non-extensive character of the system, and its proportionality to the rate of change of the multiplicity under a variation of the statistical probability parameter space, emerges naturally in the frame of this representation.

Keywords: Generalized information entropy measures, Fisher information, Tsallis, Renyi, Sharma-Mittal, entropy, Information geometry

PACS: 05.70.a, 65.40.Gr, 89.70.+c, 05.70.Ln, 05.70.a

---

[*]Corresponding author: marco_masi2@tin.it, fax: +39 02 6596997

1 Introduction

In a previous paper (18), using a formalism based on Kolmogorov-Nagumo means and generalized logarithms and exponentials, we wrote down the set of entropy functionals, from Boltzamann-Gibbs entropy through Rényi and Tsallis, up to Sharma-Mittal (25) and a new entropy measure, we called the "supra-extensive entropy", so that the increasing generalization of entropy measures from arithmetic to non-arithmetic means, and from extensive to non-extensive systems became particularly compact and visible in its hierarchical structure. Sharma-Mittal measure was already developed in 1975 but has been investigated in generalized thermostatistics only recently by Frank, Daffertshofer and Naudts ((12), (13) (20)). We showed that Sharma-Mittal's measure is however only one of two possible extensions that unify Rényi and Tsallis entropy in a coherent picture and described how it comes naturally into being together with another "supra-extensive" measure if the formalism of generalized logarithm and exponential functions is used. Moreover, we could see how the relation between these information measures is best understood when described in terms of a logarithmic Kolmogorov-Nagumo average.

In this paper we shall further investigate in particular the power of the deformed logarithm-exponential formalism with regards to the relationship of generalized entropy measures and Fisher information.

Fisher information was originally conceived in the 1920s (4), many years before Shannon's notion of entropy, as a tool of statistical inference in parameter estimation theory. It must be emphasized that Fisher's functional is an information, but not an entropy measure. There is nevertheless a strong connection between Fisher information and entropy. This relationship has been outlined in many occasions since Rao (21), already in 1945, laid the foundations of statistical differential geometry, called also *information geometry* (for a more recent review of the subject see e.g. Amari & Nagaoka (1)). Rao outlined how a statistical model can be described by a statistical differential manifold which can be considered as a Riemannian manifold of parametrized probability distributions (PD) or probability density functions (PDF) with the metric tensor given by the Fisher information matrix (FIM). The FIM determines a Riemannian information metric on this parameter space, and is therefore called also the *Fisher metric*. This has been the subject of renewed interest more recently also in other branches of information theory, in applications of image processing, econometrics and received some attention in theoretical physics, especially in regards to its, still not entirely understood, role in quantum mechanics and perhaps also quantum gravity (see e.g. B.R. Frieden's work (5) which tries to derive the laws of physics from a Fisherian point of view, or R. Carol's review (2) of some other similar attempts and references therein).

Less has been done to highlight the links between Fisher information and generalized measures and non-extensive statistics. Some attempts in this direction were made for instance by F. Pennini and A. Plastino (6), M. Portesi, F. Pennini and A. Pennini (7), S. Abe (8), J. Naudts ((9), (10)), P. Jizba (11), just to mention some examples. However, we feel that a clear exposition is lacking about the place that the Fisher information measure has in the frame of a generalized statistics. The aim of this paper is to highlight in a synthetic way the relationship that exists between Fisher information and the two-parametric generalized entropy measures here mentioned (Rényi, Tsallis, Sharma- Mittal and the supra-extensive measure which expands further the picture as a consequence of the q-deformed formalism) in the sense that diagram of page 14 illustrates, what role the two parameters play in evaluating the Fisher information matrix, and how it can be retrieved using a deformed exponential formalism. We will focus our attention on how precisely Fisher information (except a real multiplicative factor) emerges naturally as a universal statistical metric tensor for every generalized information-entropy measure defined on a manifold of PDFs (i.e. for a continuous version of the above mentioned entropies), and to obtain in a relatively simple manner this result using a representation based on generalized logarithm and exponential functions within the frame of a KN formalism.

It should also be mentioned that Rényi entropy is not Lesche-stable (14), isn't convex and does not possess the property of finite entropy production. Therefore any extension of Rényi's entropy, cannot in general possess these properties either. There is some controversy if this is supposed to have its thermodynamical implications, or not. However, the theoretical framework we are going to construct here has to be intended in a more general context, it can still have its meaning and applications in information theory, cybernetics or other fields not necessarily restricted to a generalized thermostatistics. It is with this point of view in mind that we will proceed.

2 The generalized information-entropy measures

Just to make this paper selfcontained let us briefly sum up the aspects of a generalized information-entropy measure theory which will be relevant to the understanding of the next sections.

2.1 The Boltzmann-Gibbs entropy and Shannon's information measure

As is well known the Boltzmann-Gibbs (BG) entropy reads[1]

$$S_{BG}(P) = -k \sum_i p_i \log p_i \,,$$

with p_i the probability of the system to be in the i-th microstate, k the Boltzmann constant. BG entropy becomes the celebrated Shannon information measure (24) if $k = 1$ (as we will do from now on) and uses the immaterial base b for the logarithm function (we will maintain the natural logarithm $b = e$)

$$S_S(P) = -\sum_i p_i \log_b p_i = \sum_i p_i \log_b \left(\frac{1}{p_i}\right) \equiv \sum_i p_i \log \left(\frac{1}{p_i}\right) \,. \tag{2.1}$$

BG and Shannon's measures are additive, i.e. given two systems, described by two PDs A and B, we have

$$S_S(A \cap B) = S_S(A) + S_S(B|A) \,,$$

with $S_S(B|A)$ the conditional entropy. These systems are called *extensive systems*. This is the case where the total entropy behaves as the sum of the entropies of its parts and applies to standard statistical mechanics. The additive property is reflected in the logarithm function.

2.2 Tsallis' entropy

Nature is however not always a place where additivity is preserved. This is the case of nonlinear complex systems, in fractal- or multifractal-like and self-organized critical systems, or where long range forces are at work (e.g. in star clusters or in systems with long range microscopic memory), etc. These *non-extensive systems* have been investigated especially in the last two decades (27).

Tsallis generalized Shannon's entropy to non-extensive systems as (26)

$$S_T(P, q) = \frac{\sum_i p_i^q - 1}{1 - q} = \frac{1}{q - 1} \sum_i p_i \left(1 - p_i^{q-1}\right) \,, \tag{2.2}$$

with q a real parameter. This is now widely known as *Tsallis entropy*. According to a current school of thought at least some non-extensive systems can be described by scaled power law probability functions as p_i^q, so called *q-probabilities*. For $q \to 1$ it reduces to Shannon's measure. Tsallis entropy extends additivity to *pseudo-additivity*

$$S_T(A \cap B) = S_T(A) + S_T(B|A) + (1 - q)S_T(A)S_T(B|A) \,. \tag{2.3}$$

In order to describe Tsallis sets the *generalized q-logarithm function*

$$\log_q x = \frac{x^{1-q} - 1}{1 - q} \,, \tag{2.4}$$

turns out to be particularly useful. In a similar way, its inverse, the *generalized q-exponential function* is

$$e_q^x = [1 + (1 - q)x]^{\frac{1}{1-q}} \,. \tag{2.5}$$

The classical Napier's logarithm and its inverse function is recovered for $q = 1$. The importance of the q-logarithm in this context is realized if we understand that it satisfies precisely a pseudo-additive law

$$\log_q xy = \log_q x + \log_q y + (1 - q)(\log_q x)(\log_q y) \,.$$

---

[1]Here we begin to introduce a more general symbolism according to which every type of information measure is labeled with $S_{name}(\{P\}, \{q\})$ or $\mathcal{S}_{name}(\{\mathcal{P}\}, \{q\})$, where P or $\mathcal{P}$ stands for the family $\{p_i\}$ of PDs or PDFs and S or $\mathcal{S}$ for the discrete and continuous cases respectively, while q is a scalar or vector parameter which meaning will become clear in the following sections.

Exploiting this generalized logarithm and exponential formalism Tsallis entropy 2.2 can be rewritten as

$$S_T(P, q) = -\sum_i p_i^q \log_q p_i = \sum_i p_i \log_q \left(\frac{1}{p_i}\right), \qquad (2.6)$$

which is sometimes also referred to as the *q-deformed Shannon entropy*.

Note that $\log_q x^\alpha \neq \alpha \log_q x$ when $q \neq 1$. This is the reason why, if one thinks in terms of averages, it is more meaningful to write entropy measures with the inverse of the PD, as in the r.h.s. of 2.6, and why we will prefer this formal representation.

2.3 Rényi's entropy

By looking at the structure of the r.h.s. of 2.1 and 2.6 one can define an information measure as an average of the *elementary information gains*

$$I_i \equiv I_i \left(\frac{1}{p_i}\right) = \log_q \left(\frac{1}{p_i}\right) \qquad (2.7)$$

associated to the i-th event of probability p_i

$$S_S(P) = \left\langle \log \left(\frac{1}{p_i}\right) \right\rangle_{lin} \qquad (2.8)$$

and

$$S_T(P) = \left\langle \log_q \left(\frac{1}{p_i}\right) \right\rangle_{lin} \qquad (2.9)$$

where, what is common to both, is the underlying *arithmetic-*, or *linear mean* $I = \sum_i p_i I_i$.

However, A.N. Kolmogorov and M. Nagumo ((17), (19)) showed, already in 1930 but independently from each others that, if we accept Kolomogorov's axioms as the foundation of probability theory, then the notion of average can acquire a more general meaning as what is called a *quasi-arithmetic* or *quasi-linear mean*, and can be defined as

$$S = f^{-1} \left(\sum_i p_i f(I_i)\right), \qquad (2.10)$$

with f a strictly monotonic continuous function, called the *Kolmogorov-Nagumo function* (KN-function). Rényi instead showed (22) that, if additivity is imposed on information measures, then the whole set of KN-functions must reduce to only two possible cases. The first is of course the linear mean associated with the KN-function

$$f(x) = x,$$

while the other possibility is the *exponential mean* represented by the KN-function

$$f(x) = c_1 b^{(1-q)x} + c_2; \qquad q \in \mathbb{R} \qquad (2.11)$$

with c_1 and c_2 two arbitrary constants.

Rényi's information-entropy measure is per definition a measure where the single information gains are averaged exponentially, and writes

$$S_R(P, q) = \frac{1}{1-q} \log_b \sum_i p_i^q \equiv \frac{1}{1-q} \log \sum_i p_i^q, \qquad (2.12)$$

with b the logarithm base (still we will always assume $b = e$). When $q \to 1$ Rényi's boils down to Shannon entropy.

In fact, if we choose in 2.11, $c_1 = \frac{1}{1-q} = -c_2$, then because of 2.4, it becomes

$$f(x) = log_q e^x, \qquad (2.13)$$

which inserted in 2.10 with

$$I_i = \log \left(\frac{1}{p_i}\right),$$

shows that 2.12 is equivalent to

$$S_R(P, q) = \left\langle \log \left(\frac{1}{p_i}\right) \right\rangle_{exp},$$

where $\langle \cdot \rangle_{exp}$ stands for an average defined by KN-function 2.13. Compare this with 2.8 and 2.9.

2.4 The Sharma-Mittal and *Supra-extensive* entropy

The next step in the generalization process consists in finding a measure which is non-extensive and non-additive but contains Tsallis' and Rényi's entropies as special cases. One possible way to obtain this goes through an extension of the KN-mean. This leads to what is known as the Sharma-Mittal entropy (SM) (25). However it is only by exploiting the generalized logarithm and exponential representation one retrieves in a compact and fast manner both SM entropy measure and what we used to call the "supra-extensive" (SE) entropy. By using the q-deformed logarithm and exponential formalism one could easily arrive at a further generalization of Rényi and Tsallis entropies.

The starting point is the relationship between Tsallis and Rényi entropies

$$S_R(P,q) = \frac{1}{1-q} \log \left[1 + (1-q) \, S_T(P,q) \right] .$$

From 2.4 and 2.5, we see that this is equivalent to

$$S_R(P,q) = \log e_q^{S_T(P,q)} , \qquad (2.14)$$

and therefore

$$S_T(P,q) = \log_q e^{S_R(P,q)} . \qquad (2.15)$$

2.14 and 2.15 suggest immediately two further generalization:

$$S_{SM}(P,\{q,r\}) = \log_r e_q^{S_T(P,q)} = \frac{1}{1-r} \left[\left(\sum_i p_i^q \right)^{\frac{1-r}{1-q}} - 1 \right] , \qquad (2.16)$$

and

$$S_{SE}(P,\{q,r\}) = \log_q e_r^{S_R(P,q)} = \frac{\left[1 + \frac{(1-r)}{(1-q)} \log \sum_i p_i^q \right]^{\frac{1-q}{1-r}} - 1}{1-q} , \qquad (2.17)$$

with r another real parameter.

2.16 is SM's pseudo-additive measure, while 2.17 is a new type of entropy measure we called "supra-extensive" because it generalizes to a measure which is neither additive nor pseudo-additive. We could see (18) how the decisive difference between these two information-entropies is that SM's measure can be obtained also through the KN-mean as a two parameter extension of 2.13 (with $f(x) = log_q e_r^x$ on $I_i = \log_r \left(\frac{1}{p_i} \right)$), while the SE measure does not have such kind of generalization. It can also be shown that for two systems A and B for Sharma-Mittal entropy (instead of 2.3) one has

$$S_{SM}(A \cap B) = S_{SM}(A) + S_{SM}(B|A) + (1-r)S_{SM}(A)S_{SM}(B|A) .$$

This indicates that it is the magnitude of parameter r which stands for the degree of non-extensivity, and q stands for a PD deformation parameter. When $r \to q$ the deformation parameter q of the PD merges into the non-extensivity parameter r (which is the reason why in Tsallis entropy it is q instead of r that appears for the non-extensive character of the system).

The supra-extensive entropy 2.17 however emerges naturally as a symmetric counterpart of 2.16 when generalized logarithms and exponentials are used. Further mathematical-physical investigations which will clarify the standpoint of the supra-extensive entropy, what kind of statistics it expresses, if any, and its relationship with other measures, is of course desirable and still necessary. Anyway, something can be already said. What we are going to do here is that we can show how this new entropy also shares a common status in regards to Fisher information with all the other measures too.

2.5 The multiplicity

To introduce ourselves to this, note first of all how we can rewrite the quantity

$$\left(\sum_i p_i^q \right)^{\frac{1}{1-q}} = \left(\sum_i p_i \left(\frac{1}{p_i} \right)^{1-q} \right)^{\frac{1}{1-q}} = \left\langle \left(\frac{1}{p_i} \right)^{1-q} \right\rangle_{\text{lin}}^{\frac{1}{1-q}} = e_q^{\left\langle log_q \left(\frac{1}{p_i} \right) \right\rangle_{\text{lin}}}$$

$$= e_q^{S_T(P,q)} = \left\langle \frac{1}{p_i} \right\rangle_{\log_q} \equiv \Omega(P,q)\,, \tag{2.18}$$

where we used what we call the *logarithmic mean* $\langle \cdot \rangle_{\log_q}$ defined by the KN-function $f(x) = \log_q x$.

Then using 2.5, equations 2.14 to 2.17 can be rewritten as

$$S_T(P,q) = \log_q \left\langle \frac{1}{p_i} \right\rangle_{\log_q} = \log_q \Omega(P,q)\,; \tag{2.19}$$

$$S_R(P,q) = \log \left\langle \frac{1}{p_i} \right\rangle_{\log_q} = \log \Omega(P,q)\,; \tag{2.20}$$

$$S_{SM}(P,\{q,r\}) = \log_r \left\langle \frac{1}{p_i} \right\rangle_{\log_q} = \log_r \Omega(P,q)\,; \tag{2.21}$$

$$S_{SE}(P,\{q,r\}) = \log_q e_r^{\log \left\langle \frac{1}{p_i} \right\rangle_{\log_q}} = \log_q e_r^{\log \Omega(P,q)}\,. \tag{2.22}$$

Rewriting things in the language of this representation and using the KN logarithmic mean one can see more straightforwardly how Sharma-Mittal's entropy generalizes Rényi's extensive entropy to non-extensivity, and how the new measure does the same for non-extensivity generalizing it to a 'generalized non-extensivity', we called *supra-extensivity*.

The quantity

$$\Omega(P,q) = \left\langle \frac{1}{p_i} \right\rangle_{\log_q} = \left(\sum_i p_i^q \right)^{\frac{1}{1-q}}\,,$$

is well known to have a physical interpretation in statistical mechanics: the multiplicity of the system, i.e. the number of all possible microstates compatible with its macroscopic state.

3 Generalizing to relative entropy-information measures

S. Kullback and R. A. Leibler (15) introduced the notion of *relative entropy*.

Given a random variable X with $\mathbf{x}$ a specific (scalar or vector) value for X on a continuous event space, consider continuous differentiable PDFs, $p(x,\theta) \in \mathcal{C}^2$, with θ a (scalar or vector) parameter. Let be H_1 the hypotheses that X is from the statistical population with PDF $p_1(\mathbf{x},\theta)$ and H_2 that with PDF $p_2(\mathbf{x},\phi)$. Then it can be shown (16) that applying Bayes' theorem, $\log \frac{p_1(\mathbf{x},\theta)}{p_2(\mathbf{x},\phi)}$ measures the difference between the *logarithm of the odds in favor of H_1 against H_2* before a measurement gave $X = \mathbf{x}$. *Kullback's relative entropy*, or our "mean capacity for discrimination" in favor of H_1 against H_2, was originally defined as

$$\mathcal{E}_{KL} = \mathcal{E}_{KL}(\{p_1,p_2\}) = \int_{\mathcal{S}_{sp}} p_1(\mathbf{x},\theta) \log \frac{p_1(\mathbf{x},\theta)}{p_2(\mathbf{x},\phi)} \, d^n\mathbf{x}\,,$$

with $\mathcal{S}_{sp}$ the entire sample space.

If $p_2(\mathbf{x},\phi) = 1$ (we "discriminate" against certainty), the negative Shannon information (in its continuous form) is recovered. The different signature is due to the fact that Shannon's information, as all the measures we are dealing with here, account for the amount of information we still need to gain complete knowledge, i.e. the uncertainty about the message. Let us therefore call *Kullback-Leibler relative information-entropy measure*, or simply *Kullback's measure*

$$\mathcal{S}_{KL} = \mathcal{S}_{KL}(\{p_1,p_2\}) = \int_{\mathcal{S}_{sp}} p_1(\mathbf{x},\theta) \log \frac{p_2(\mathbf{x},\phi)}{p_1(\mathbf{x},\theta)} \, d^n\mathbf{x}\,. \tag{3.1}$$

Relative entropies can be used to generalize all information measures either in their continuous as in their discrete version. Let us start first with discrete PDs.

Given two families of PDs $P = \{P^{(1)}; P^{(2)}\} = \{p_i^{(1)}; p_j^2\}$, $(i,j = (1,...,\Omega))$, Kullback's measure 3.1 takes the form

$$S_{KL}(P) = \sum_i p_i^{(1)} \log \left(\frac{p_i^{(2)}}{p_i^{(1)}} \right) = \left\langle \log \left(\frac{p_i^{(2)}}{p_i^{(1)}} \right) \right\rangle_{\mathrm{lin}} = \log \left\langle \left(\frac{p_i^{(2)}}{p_i^{(1)}} \right) \right\rangle_{\log}\,. \tag{3.2}$$

Then, in a more general context, we can extend 2.7 to *elementary relative information gains* as

$$I_i = \log\left(\frac{p_i^{(2)}}{p_i^{(1)}}\right) \qquad \text{(for extensive systems)},$$

or

$$I_i = \log_s\left(\frac{p_i^{(2)}}{p_i^{(1)}}\right) \qquad \text{(for non} - \text{extensive systems)},$$

in

$$I = f^{-1}\left(\sum_i p_i^{(1)} f(I_i)\right),$$

with $s = q$ or $s = r$ for Tsallis' and SM's entropies respectively, that is we can rewrite 2.19-2.21 with all KN means so far considered again generalizing it to relative information gains, and then replace the so obtained relative Rényi entropy in the exponential expression of 2.22 (or, proceeding in a somewhat less rigorous manner, simply extend $\frac{1}{p_i} \to \frac{p_i^{(2)}}{p_i^{(1)}}$ in all of them)

$$S_T(P,q) = \log_q \left\langle \frac{p_i^{(2)}}{p_i^{(1)}} \right\rangle_{\log_q} = \frac{1}{1-q}\left[\sum_i (p_i^{(1)})^q (p_i^{(2)})^{1-q} - 1\right], \tag{3.3}$$

$$S_R(P,q) = \log \left\langle \frac{p_i^{(2)}}{p_i^{(1)}} \right\rangle_{\log_q} = \frac{1}{1-q} \log \sum_i (p_i^{(1)})^q (p_i^{(2)})^{1-q}, \tag{3.4}$$

$$S_{SM}(P,\{q,r\}) = \log_r \left\langle \frac{p_i^{(2)}}{p_i^{(1)}} \right\rangle_{\log_q} = \frac{1}{1-r}\left[\left(\sum_i (p_i^{(1)})^q (p_i^{(2)})^{1-q}\right)^{\frac{1-r}{1-q}} - 1\right], \tag{3.5}$$

$$S_{SE}(P,\{q,r\}) = \log_q e_r^{\log\left\langle \frac{p_i^{(2)}}{p_i^{(1)}} \right\rangle_{\log_q}} = \frac{\left[1 + \frac{(1-r)}{(1-q)} \log \sum_i (p_i^{(1)})^q (p_i^{(2)})^{1-q}\right]^{\frac{1-q}{1-r}} - 1}{1-q}. \tag{3.6}$$

For $p_i^{(2)} \to 1$ they reduce to 2.2, 2.12, 2.16 and 2.17 respectively, while for $q = 1$ Tsallis' and Rényi's measures 3.3 and 3.4 become both Kullback's measure 3.2. From 3.5 (3.6) we recover Rényi's (Tsallis') measure 3.4 (3.3), if $r \to 1$, and Tsallis (Rényi's) measure 3.3 (3.4), if $r \to q$. Notice how it is much easier to recognize the limits in the logarithmic-exponential representation.

Straightforwardly we can now extend to continuous PDFs over parameter spaces θ and ϕ. The *continuous Tsallis, Rényi, Sharma-Mittal and supra-extensive relative information-entropy measures* become

$$\mathcal{S}_T(\mathcal{P},q) = \frac{1}{1-q}\left[\int p_1(\mathbf{x},\theta)^q p_2(\mathbf{x},\phi)^{1-q}\, d^n\mathbf{x} - 1\right]; \tag{3.7}$$

$$\mathcal{S}_R(\mathcal{P},q) = \frac{1}{1-q} \log \int p_1(\mathbf{x},\theta)^q p_2(\mathbf{x},\phi)^{1-q}\, d^n\mathbf{x}; \tag{3.8}$$

$$\mathcal{S}_{SM}(\mathcal{P},\{q,r\}) = \frac{1}{1-r}\left[\left(\int p_1(\mathbf{x},\theta)^q p_2(\mathbf{x},\phi)^{1-q}\, d^n\mathbf{x}\right)^{\frac{1-r}{1-q}} - 1\right]; \tag{3.9}$$

$$\mathcal{S}_{SE}(\mathcal{P},\{q,r\}) = \frac{\left[1 + \frac{(1-r)}{(1-q)} \log \int (p_1(\mathbf{x},\theta))^q (p_2(\mathbf{x},\phi))^{1-q}\right]^{\frac{1-q}{1-r}} - 1}{1-q}. \tag{3.10}$$

Of course, one could again rewrite things all over again, to see that the same result appears if we extend the Kolmogorov-Nagumo mean to continuity as

$$\mathcal{S} = f^{-1}\left(\int p_1(\mathbf{x},\theta) f(\mathcal{I}_x(\mathbf{x},\theta,\phi))\, d^n\mathbf{x}\right), \tag{3.11}$$

where

$$\mathcal{I}_x(\mathbf{x}, \theta, \phi) = \log\left(\frac{p_2(\mathbf{x}, \phi)}{p_1(\mathbf{x}, \theta)}\right) \qquad \text{for extensive systems}\,;$$

or

$$\mathcal{I}_x(\mathbf{x}, \theta, \phi) = \log_q\left(\frac{p_2(\mathbf{x}, \phi)}{p_1(\mathbf{x}, \theta)}\right) \qquad \text{for non-extensive systems}\,,$$

and/or using the generalized q-deformed logarithm and exponential expressions from 2.19 to 2.22, extending $\frac{1}{p_i} \to \frac{p_2(\mathbf{x}, \phi)}{p_1(\mathbf{x}, \theta)}$.

Then, applying 3.11 ($f = \log_q x$) to obtain the relative and continuous extension of multiplicity 2.5, one has [2]

$$\Omega(\theta, \phi) = \left\langle \frac{p_2(x, \phi)}{p_1(x, \theta)} \right\rangle_{\log_q} = e_q^{\int p_1(x,\theta) \log_q\left(\frac{p_2(x,\phi)}{p_1(x,\theta)}\right) d^n\mathbf{x}} = e_q^{S_T(\theta, \phi)}\,. \tag{3.12}$$

Then we can rewrite 3.7 to 3.10 in its relative continuous extension of 2.19 to 2.22 as

$$\mathcal{S}_T(\theta, \phi) = \log_q\left\langle \frac{p_2(x, \phi)}{p_1(x, \theta)} \right\rangle_{\log_q} = \log_q \Omega(\theta, \phi)\,; \tag{3.13}$$

$$\mathcal{S}_R(\theta, \phi) = \log\left\langle \frac{p_2(x, \phi)}{p_1(x, \theta)} \right\rangle_{\log_q} = \log \Omega(\theta, \phi)\,; \tag{3.14}$$

$$\mathcal{S}_{SM}(\theta, \phi) = \log_r\left\langle \frac{p_2(x, \phi)}{p_1(x, \theta)} \right\rangle_{\log_q} = \log_r \Omega(\theta, \phi)\,; \tag{3.15}$$

$$\mathcal{S}_{SE}(\theta, \phi) = \log_q e_r^{\log\left\langle \frac{p_2(x, \phi)}{p_1(x, \theta)} \right\rangle_{\log_q}} = \log_q e_r^{\log \Omega(\theta, \phi)}\,. \tag{3.16}$$

4 The role of Fisher information for generalized entropy measures

4.1 The Fisher information measure

We are now ready to proceed towards the real aim of this paper. We begin with a brief introduction to Fisher information.

In 1921, R. Arnold Fisher defined an information measure which could account for the "quality" or "efficiency" of a measurement. Calling *efficient estimator* or *best estimator*, the best unbiased estimate $\widehat{\theta}(x)$ of θ after many independent measurements on a random variable x such that $<\widehat{\theta}(x)> = \int p(x, \theta)\,\widehat{\theta}(x)\,dx = \theta$, Fisher defined the *efficiency* or *quality* of a measurement, $\mathcal{I}_F$, the quantity which satisfies

$$\mathcal{I}_F\, e^2 = 1\,,$$

where $e^2 = \int p(x, \theta)\,[\widehat{\theta}(x) - \theta]^2\,dx$ is the mean square error.

Fisher showed (4) that then $\mathcal{I}_F$ is uniquely identified as

$$\mathcal{I}_F(\mathcal{P}) = \left\langle \left(\frac{\partial \log p(x, \theta)}{\partial \theta}\right)^2 \right\rangle_{lin} =$$

$$= \int_{\mathcal{S}_{sp}} p(x, \theta)\left(\frac{\partial \log p(x, \theta)}{\partial \theta}\right)^2 dx = \int_{\mathcal{S}_{sp}} \frac{1}{p(x, \theta)}\left(\frac{\partial p(x, \theta)}{\partial \theta}\right)^2 dx\,.$$

For any other estimator one chooses, the *Cramer-Rao inequality*, or *Cramer-Rao bound*, holds

$$\mathcal{I}_F\, e^2 \geq 1\,.$$

---

[2]Since we will work with parameters, let us write for a lighter notation on the multiplicity and the entropies, $\Omega(\mathcal{P}, \{q\}) \equiv \Omega(\theta, \phi)$ and $\mathcal{S}(\mathcal{P}, \{q, r\}) \equiv \mathcal{S}(\theta, \phi)$.

Going over to N-dimensional vector random variables $\mathbf{x} = (x_1, ..., x_N)$ on an M-dimensional parameter space $\theta = (\theta_1, ..., \theta_M)$, Fisher defined its celebrated (symmetric) *Fisher information matrix* which elements are given by

$$
\begin{aligned}
F_{ij}(\theta) &= \left\langle \frac{\partial \log p(\mathbf{x}, \theta)}{\partial \theta_i} \frac{\partial \log p(\mathbf{x}, \theta)}{\partial \theta_j} \right\rangle_{lin} \\
&= \int_{\mathcal{S}_{sp}} p(\mathbf{x}, \theta) \frac{\partial \log p(\mathbf{x}, \theta)}{\partial \theta_i} \frac{\partial \log p(\mathbf{x}, \theta)}{\partial \theta_j} d^n\mathbf{x} \\
&= \int_{\mathcal{S}_{sp}} \frac{1}{p(\mathbf{x}, \theta)} \frac{\partial p(\mathbf{x}, \theta)}{\partial \theta_i} \frac{\partial p(\mathbf{x}, \theta)}{\partial \theta_j} d^n\mathbf{x} \,,
\end{aligned} \tag{4.1}
$$

with $(i, j = 1, .., M)$. If we would further extend to an L-dimensional continuous probability space $\mathcal{P} = (p_1, ..., p_L)$, then the most general expression for Fisher information writes

$$
\mathcal{I}_F(\mathcal{P}) = \sum_{k=1}^{L} \sum_{i,j=1}^{M} F_{ij}^k(\theta) \,.
$$

4.2 The Fisher information matrix as a metric tensor

We will not go into the details in what would be a much too long exposition of information geometry and shall highlight only in an introductory manner the status of the FIM as a metric tensor for a statistical manifold (for a more rigorous account of the subject see e.g. (1), (3), (28), (23), and references therein).

Consider a family of $\mathcal{C}^2$ differentiable PDFs with N-dimensional continuous vector random variables $\mathbf{x}$, parametrized by an M-dimensional continuous real vector parameter space θ on an open interval $I_\theta \subseteq \mathbb{R}^M$

$$
\mathcal{F}_\theta = \{p(\mathbf{x}, \theta) \in \mathcal{C}^2; \theta \in I_\theta\} \,.
$$

The notion of a *differential statistical manifold* is identified in the fact that the parameters θ can be conceived as providing a local coordinate system for an M-dimensional manifold $\mathcal{M}$ which points are in a one to one correspondence with the distributions $p \in \mathcal{F}_\theta$.

Since information-entropy measures are log-probability functionals defined on $\mathcal{M}$, it is convenient to consider also the function $l : \mathcal{M} \to \mathbb{R}$ on the manifold $\mathcal{M}$ defined as $l(\theta) \equiv \log p(\mathbf{x}, \theta)$. This is commonly called the *log-likelihood function*. Labelling the manifold's tangent space $T_\theta(\mathcal{M})$, the directional derivatives of $l(\theta)$ along the tangent vectors $\widehat{\mathbf{e}}_i \in T_\theta(\mathcal{M})$ at a point in $\mathcal{M}$ with coordinates θ are (use the shorthand $\partial_i \equiv \frac{\partial}{\partial \theta_i}$): $\partial_i l(\theta) \widehat{\mathbf{e}}_i = \frac{\partial_i p(\mathbf{x}, \theta)}{p(\mathbf{x}, \theta)} \widehat{\mathbf{e}}_i$.

FIM 4.1 can also be seen as the expectation value with respect to $p(\mathbf{x}, \theta)$ of the partial derivatives of $l(\theta)$, which is the reason why in the literature it is frequently written as

$$
F_{ij}(\theta) = E[\partial_i l(\theta) \, \partial_j l(\theta)] \,.
$$

This is a symmetric, non-degenerate, bilinear form on a vector space of random variables $\partial_i l(\theta)$. But a Riemannian metric g is per definition a symmetric non-degenerate inner product on the manifold's tangent space $T_\theta(\mathcal{M})$, and one can therefore consider the FIM as the statistical analogue of the metric tensor for a statistical manifold.

By the way, it is worth mentioning that Corcuera & Giummol'e showed (3) that the FIM has also the unique properties of being covariant under reparametrization of the parameter space of the manifold, and invariant under reparametrization of the sample space (see also Wagenaar (28) for a review). This is an appealing aspect which possibly suggests that Fisher information might play some role in future quantum spacetime theories.

Now, the metric tensor tells how to compute the distance between any two points in a given space. Here we are considering the distance between two points on a statistical differential manifold mapped on a measure functional, i.e. the informational difference between them. This idea can be introduced regarding Kullback's relative information measure to account for the net dissimilarity between two families of PDFs with parameters, θ and ϕ. Intuitively one can imagine this as measuring a "distance" between these two families. However, strictly speaking, this is not a metric distance because it is neither symmetric nor satisfies the triangle inequality (on statistical manifolds one has to consider an extended version of

Pythagora's law). The symmetry condition however can be restored if instead of the single information measure we use the *divergence* $\mathcal{D}$ of two PDFs, p_1 and p_2, defined as[3]

$$\mathcal{D}(p_1, p_2) = \frac{\mathcal{S}(p_1, p_2) + \mathcal{S}(p_2, p_1)}{2} \, .$$

If we choose to set

$$p_1(\mathbf{x}, \theta) \equiv p(\mathbf{x}, \theta); \quad p_2(\mathbf{x}, \phi) \equiv p(\mathbf{x}, \theta + d\theta), \tag{4.2}$$

then the symmetric divergence $\mathcal{D}(p(\mathbf{x}, \theta), p(\mathbf{x}, \theta + d\theta)) \equiv \mathcal{D}(\theta, \theta + d\theta)$ can be intended as an extension of the square of the Riemannian distance between two nearby distributions. Expanded to second order it gives

$$\mathcal{D}(\theta, \theta + d\theta) = \frac{1}{2!} \sum_{ij} \left[\frac{\partial^2 \mathcal{D}(\theta, \phi)}{\partial \theta_i \partial \theta_j} \right]_{\phi = \theta} d\theta_i d\theta_j \, + \, O(d\theta^3) \, ,$$

because $\mathcal{D}(\theta, \phi)$ is minimal at $\phi = \theta$ and the first order vanishes. It is the second order, not the first, which is the leading one in every information measure divergence, and it can be shown ((1), (3), (28)) that it is the second derivative of the divergence which defines the metric, i.e.

$$g_{ij}(\theta) = \left[\frac{\partial^2 \mathcal{D}(\theta, \phi)}{\partial \theta_i \partial \theta_j} \right]_{\phi = \theta} = \frac{1}{2} \left[\frac{\partial^2 (\mathcal{S}(\theta, \phi) + \mathcal{S}(\phi, \theta))}{\partial \theta_i \partial \theta_j} \right]_{\phi = \theta} \, . \tag{4.3}$$

In case of Kullback's measure 3.1, the divergence is defined as

$$\mathcal{D}_{KL}(\theta, \phi) = \frac{1}{2} \int [p(\mathbf{x}, \theta) - p(\mathbf{x}, \phi)] \log \frac{p(\mathbf{x}, \phi)}{p(\mathbf{x}, \theta)} \, d^n\mathbf{x} \, .$$

From 4.3, and keeping in mind that if we want the normalization condition to hold for every θ implies

$$\frac{\partial}{\partial \theta_i} 1 = \frac{\partial}{\partial \theta_i} \int p(\mathbf{x}, \theta) d^n\mathbf{x} = \int \frac{\partial}{\partial \theta_i} p(\mathbf{x}, \theta) d^n\mathbf{x} = \int \frac{\partial^2}{\partial \theta_i \theta_j} p(\mathbf{x}, \theta) d^n\mathbf{x} = 0 \, , \tag{4.4}$$

we have

$$g_{ij}^{KL}(\theta) = - \int \frac{1}{p(\mathbf{x}, \theta)} \frac{\partial p(\mathbf{x}, \theta)}{\partial \theta_i} \frac{\partial p(\mathbf{x}, \theta)}{\partial \theta_j} \, d^n\mathbf{x} = -F_{ij}(\theta) \, ,$$

which is the (i,j)-th element of the negative FIM 4.1.

This is a very important and known result from information geometry. It is in this sense that g_{ij} can be seen as a metric tensor which measures a "distance" on a statistical manifold in a Riemannian space. In this sense Fisher information can be said to be a sort of "mother information measure".

5 The Fisher metric for generalized information-entropy measures

We can generalize this result of information geometry. The Fisher metric for Tsallis, Rényi, the Sharma-Mittal and the supra-extensive measures can be obtained considering the relative entropy measures as defined in 3.7, 3.8, 3.9 and 3.10 respectively (with $p_1 = p(\mathbf{x}, \theta)$, $p_2 = p(\mathbf{x}, \phi)$), from their respective symmetric *divergence*

$$\mathcal{D}(\theta, \phi) = \frac{\mathcal{S}(\theta, \phi) + \mathcal{S}(\phi, \theta)}{2} \, ,$$

defined on $\mathcal{F}_\theta$.

What we need is the evaluation of 4.3 for each information measure 3.7 to 3.10. One can of course compute directly the (somewhat fuzzy) second derivatives $\frac{\partial^2 \mathcal{S}(\phi, \theta)}{\partial \theta_i \partial \theta_i}$ each time (and for each $\theta \to \phi$ parameter exchange). However, the q-deformed generalized logarithm and exponential formalism and the KN-representation make this task easier since it needs only the evaluation of Tsallis' entropy, the rest follows almost automatically.

The final result will be that g_{ij}^{KL} remains still the fundamental quantity, but for these more general (supra-extensive, Sharma-Mittal, Rényi and Tsallis) relative entropies the statistical metric tensor

---

[3]The notion of divergence in information geometry can be established in a rigorous way and is much more general. We shall however use only this particular type of definition because it is sufficient for our purposes.

$(g_{ij}^{SE}, g_{ij}^{SM}, g_{ij}^{R}$ and g_{ij}^{T} respectively) turns out to be only slightly extended by a scalar multiplicative q-deforming factor as

$$g_{ij}^{SE}(\theta) = g_{ij}^{SM}(\theta) = g_{ij}^{R}(\theta) = g_{ij}^{T}(\theta) = q\, g_{ij}^{KL}(\theta) = -q\, F_{ij}(\theta)\,. \tag{5.1}$$

This shows also that while g_{ij} depends from the q-deforming parameter it is independent from the r-extensivity parameter. This is quite natural since Fisher information accounts for the "quality" of a measure, or so to say, our "differential capacity to distinguish" locally between two neighboring PDFs, and this in turn depends from the "form" of the PDF (the q-scaling), but is independent from the extensive, non-extensive or supra-extensive character, since these are global features of the system. We shall see how it is the normalization condition imposed on PDFs that leads to this independency (and recover the known fact that this is also the same reason why g_{ij} is symmetric). Moreover, it will also become clear how Fisher information measures the rate of change of the multiplicity under a parameter variation.

5.1   Fisher from Tsallis information

First of all consider the derivation rules

$$\frac{\partial \log_q x}{\partial x} = \frac{1}{x^q}\,; \qquad \frac{\partial e_q^x}{\partial x} = \left(e_q^x\right)^q\,. \tag{5.2}$$

Writing Tsallis' continuous relative entropy 3.7 in the q-deformed Shannon notation of 2.6, we have

$$\mathcal{S}_T(\theta, \phi) = \int p_1(\mathbf{x}, \theta) \log_q \left(\frac{p_2(\mathbf{x}, \phi)}{p_1(\mathbf{x}, \theta)}\right) d^n \mathbf{x}\,. \tag{5.3}$$

We must be careful in remembering that in general the entropy measures considered are not symmetric and have to consider also

$$\mathcal{S}_T(\phi, \theta) = \int p_2(\mathbf{x}, \phi) \log_q \left(\frac{p_1(\mathbf{x}, \theta)}{p_2(\mathbf{x}, \phi)}\right) d^n \mathbf{x}\,.$$

Then, applying the q-logarithm derivation rule 5.2, one obtains for the first case (as before $\frac{\partial}{\partial \theta_i} \equiv \partial_i$; $\frac{\partial^2}{\partial \theta_i \partial \theta_j} \equiv \partial_{ij}$)

$$\partial_i \mathcal{S}_T(\theta, \phi) = \int \left[\log_q \left(\frac{p_2(\mathbf{x}, \phi)}{p_1(\mathbf{x}, \theta)}\right) - \left(\frac{p_2(\mathbf{x}, \phi)}{p_1(\mathbf{x}, \theta)}\right)^{1-q}\right] \partial_i p_1(\mathbf{x}, \theta)\, d^n \mathbf{x}\,, \tag{5.4}$$

and

$$\partial_{ij} \mathcal{S}_T(\theta, \phi) = -q \int \left(\frac{p_2(\mathbf{x}, \phi)}{p_1(\mathbf{x}, \theta)}\right)^{1-q} \frac{1}{p_1(\mathbf{x}, \theta)}\, \partial_i p_1(\mathbf{x}, \theta) \partial_j p_1(\mathbf{x}, \theta)\, d^n \mathbf{x}$$

$$+ \int \left[\log_q \left(\frac{p_2(\mathbf{x}, \phi)}{p_1(\mathbf{x}, \theta)}\right) - \left(\frac{p_2(\mathbf{x}, \phi)}{p_1(\mathbf{x}, \theta)}\right)^{1-q}\right] \partial_{ij} p_1(\mathbf{x}, \theta) d^n \mathbf{x}\,.$$

While in the second case one has quite different derivatives

$$\partial_i \mathcal{S}_T(\phi, \theta) = \int \left(\frac{p_2(\mathbf{x}, \phi)}{p_1(\mathbf{x}, \theta)}\right)^q \partial_i p_1(\mathbf{x}, \theta)\, d^n \mathbf{x}\,, \tag{5.5}$$

and

$$\partial_{ij} \mathcal{S}_T(\phi, \theta) = -q \int \left(\frac{p_2(\mathbf{x}, \phi)}{p_1(\mathbf{x}, \theta)}\right)^q \frac{1}{p_1(\mathbf{x}, \theta)}\, \partial_i p_1(\mathbf{x}, \theta) \partial_j p_1(\mathbf{x}, \theta)\, d^n \mathbf{x} +$$

$$+ \int \left(\frac{p_2(\mathbf{x}, \phi)}{p_1(\mathbf{x}, \theta)}\right)^q \partial_{ij} p_1(\mathbf{x}, \theta)\, d^n \mathbf{x}\,.$$

Note that these derivatives are not the same that one would obtain directly from 3.7, because in that case one assumes implicitly the normalization condition satisfied a priori. 3.7 and 5.3 are numerically identical only for a normalized PDF. The logarithmic-exponential representation, as in the latter case, does therefore not only represent a more general expression but, highlights better where and with

what effects the normalization enters into the play. Restricting to PDFs as 4.2, then, because of the normalization condition 4.4, from 5.4 and 5.5 one obtains

$$[\partial_i \mathcal{S}_T(\theta, \phi)]_{\phi=\theta} = [\partial_i \mathcal{S}_T(\phi, \theta)]_{\phi=\theta} = 0 \,, \tag{5.6}$$

while remembering the expression for the FIM 4.1

$$[\partial_{ij} \mathcal{S}_T(\theta, \phi)]_{\phi=\theta} = [\partial_{ij} \mathcal{S}_T(\phi, \theta)]_{\phi=\theta} =$$

$$= -q \int \frac{1}{p(\mathbf{x}, \theta)} \, \partial_i p(\mathbf{x}, \theta) \partial_j p(\mathbf{x}, \theta) \, d^n \mathbf{x} = -q F_{ij}(\theta) \,,$$

which, through 4.3, gives us finally $g_{ij}^T = -q F_{ij}(\theta)$.

So, since the FIM is symmetric, by the way, we see that in this case, and as we shall see also in all the others, it is in particular the normalization condition which renders the statistical metric tensor g_{ij} symmetric.

5.2   Fisher from Rényi information

Evaluating Tsallis' derivatives is indispensable but, once established, we don't need to make any direct derivative anymore for all the other measures if we work with generalized logarithms and exponentials. We don't even need to repeat the derivation for the symmetry considerations.

In fact, 4.3 for Rényi's measure can be obtained from 2.14. From 5.2 we obtain (the arguments $(\mathbf{x}, \theta)$ or $(\mathbf{x}, \phi)$ of the measures, the PDFs or of the FIM, shall be omitted if it is not needed otherwise)

$$\partial_i \mathcal{S}_R = \partial_i \log e_q^{\mathcal{S}_T} = \left(e_q^{\mathcal{S}_T} \right)^{q-1} \partial_i \mathcal{S}_T \,, \tag{5.7}$$

and

$$\partial_{ij} \mathcal{S}_R = \left(e_q^{\mathcal{S}_T} \right)^{q-1} \left[\partial_{ij} \mathcal{S}_T + (q-1) \left(e_q^{\mathcal{S}_T} \right)^{q-1} \partial_i \mathcal{S}_T \partial_j \mathcal{S}_T \right] \,.$$

Since $[\mathcal{S}_T]_{\phi=\theta} = 0$, applying the normalization condition (i.e. because of 5.6), we have

$$[\partial_{ij} \mathcal{S}_R]_{\phi=\theta} = [\partial_{ij} \mathcal{S}_T]_{\phi=\theta} \,, \tag{5.8}$$

which leads us to state $g_{ij}^R = g_{ij}^T = -q F_{ij}$.

5.3   Fisher from Sharma-Mittal information

Use Sharma-Mittal entropy as given in 2.16 and proceed as in the previous case

$$\partial_i \mathcal{S}_{SM} = \partial_i \log_r e_q^{\mathcal{S}_T} = \left(e_q^{\mathcal{S}_T} \right)^{q-r} \partial_i \mathcal{S}_T \,,$$

and

$$\partial_{ij} \mathcal{S}_{SM} = \left(e_q^{\mathcal{S}_T} \right)^{q-r} \left[\partial_{ij} \mathcal{S}_T + (q-r) \left(e_q^{\mathcal{S}_T} \right)^{q-1} \partial_i \mathcal{S}_T \, \partial_j \mathcal{S}_T \right] \,.$$

And again because of 5.6

$$[\partial_{ij} \mathcal{S}_{SM}]_{\phi=\theta} = [\partial_{ij} \mathcal{S}_T]_{\phi=\theta} \,,$$

we have again $g_{ij}^{SM} = g_{ij}^T = -q F_{ij}$. Note that it is the normalization condition, forcing the r.h.s. derivatives to vanish, which leads to the independency of g_{ij} from the non-extensivity parameter r.

5.4   Fisher from supra-extensive information

From 2.17 we get

$$\partial_i \mathcal{S}_{SE} = \partial_i \log_q e_r^{\mathcal{S}_R} = \left(e_r^{\mathcal{S}_R} \right)^{r-q} \partial_i \mathcal{S}_R \,,$$

and

$$\partial_{ij} \mathcal{S}_{SE} = \left(e_r^{\mathcal{S}_R} \right)^{r-q} \left[\partial_{ij} \mathcal{S}_R + (r-q) \left(e_r^{\mathcal{S}_R} \right)^{r-1} \partial_i \mathcal{S}_R \, \partial_j \mathcal{S}_R \right] \,.$$

Because of 5.6 and 5.7

$$[\partial_i \mathcal{S}_R(\theta, \phi)]_{\phi=\theta} = [\partial_i \mathcal{S}_R(\phi, \theta)]_{\phi=\theta} = 0 \,,$$

then, remembering 5.8 one has $[\partial_{ij}\mathcal{S}_{SE}]_{\phi=\theta} = [\partial_{ij}\mathcal{S}_R]_{\phi=\theta} = [\partial_{ij}\mathcal{S}_T]_{\phi=\theta}$, and finally

$$g_{ij}^{SE} = g_{ij}^T = -qF_{ij}\,.$$

Therefore, either $\partial_{ij}\mathcal{S}_{SM}$ as $\partial_{ij}\mathcal{S}_{SE}$ don't depend from the r parameter because of the normalization condition.

5.5 Working with the multiplicity

Just for didactics, in order to show how the generalized exponential-logarithmic formalism combined with the KN expressions can be used, we reach the same conclusion from the perspective of the entropies as a measure of multiplicity. From 3.12 one has

$$\partial_i\Omega = \left(e_q^{S_T}\right)^q \partial_i S_T = \Omega^q\,\partial_i S_T \;\; \Rightarrow\; [\partial_i\Omega(\theta,\phi)]_{\phi=\theta} = [\partial_i\Omega(\phi,\theta)]_{\phi=\theta} = 0\,,$$

while

$$\partial_{ij}\Omega = \Omega^q\big[\partial_{ij}S_T + q\,\Omega^{q-1}\partial_i S_T\partial_j S_T\big]\,,$$

which implies that

$$[\partial_{ij}\Omega]_{\phi=\theta} = [\partial_{ij}S_T]_{\phi=\theta} = -qF_{ij}\,.$$

Therefore, working with information-entropy measures expressed with the multiplicity as in 3.13 to 3.16, the SM's measure second derivative is

$$\partial_{ij}S_{SM} = \frac{\partial_{ij}\Omega}{\Omega^r} - r\frac{\partial_i\Omega\,\partial_j\Omega}{\Omega^{r+1}}\,,$$

one has

$$[\partial_{ij}S_{SM}]_{\phi=\theta} = -qF_{ij}\,,$$

and the above results for Tsallis, Rényi and Shannon's measure all follow again as special cases.

Finally, for SE measure

$$\partial_{ij}S_{SE} = \partial_j\left(\left(e_r^{\log\Omega}\right)^{r-q}\frac{\partial_i\Omega}{\Omega}\right)$$

$$= (r-q)\left(e_r^{\log\Omega}\right)^{2r-q-1}\frac{\partial_i\Omega}{\Omega}$$

$$+ \left(e_r^{\log\Omega}\right)^{r-q}\left(\frac{\partial_{ij}\Omega}{\Omega} - \frac{\partial_i\Omega\,\partial_j\Omega}{\Omega^2}\right)\,,$$

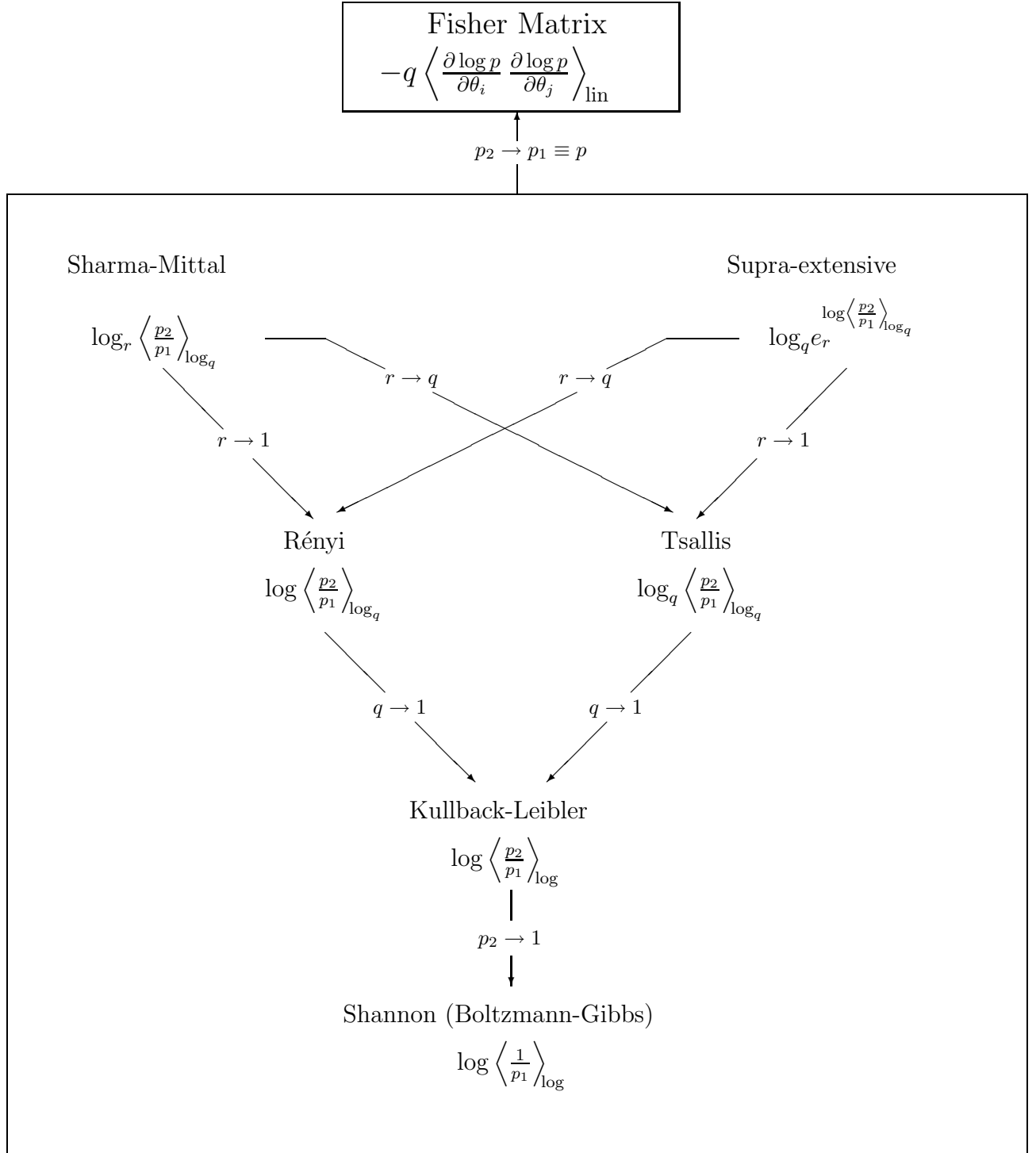and, as was to expect, the final results simplifies to

$$[\partial_{ij}S_{SE}]_{\phi=\theta} = [\partial_{ij}\Omega]_{\phi=\theta} = -qF_{ij}\,.$$

Therefore, since the second order of the multiplicity is the leading one, we can say that Fisher information accounts (times a negative parameter multiplicative deformation factor) for the change of multiplicity (the change of number of microstates of a system) under a statistical parameter variation. This is another way to interpret the fundamental connection between Fisher information and entropy measures.

6 Conclusion

Using the notion of Kullback-Leibler's relative entropy, generalizing it to all entropies, we showed, as it was already known for Kullback's measure, that once again the FIM appears as the same statistical metric tensor 5.1 for Tsallis, Rényi, Sharma-Mittal and the supra-extensive measures too. The differential-geometric properties of the divergence for each measure are independent from the extensive, non-extensive or supra-extensive character of the system, but depend only from the q-deforming parameter. This independency and the symmetry of g_{ij} are guaranteed by the normalization condition. We could also see how Fisher information has to be interpreted as a quantity proportional to the change of the information multiplicity under the statistical parameter variation. Generally, the derivation of Fisher information proved to be easier to obtain by exploiting the q-deformed logarithm and exponential formalism or the KN-representation of information-entropy measures. The overall global picture of the generalization process we have undertaken so far can be finally summarized in the diagram of the following page (where p_1 and p_2 can be both PDs or PDFs).

Hierarchy of generalized relative entropy measures

$$\boxed{\begin{array}{c} \text{Fisher Matrix} \\ -q \left\langle \dfrac{\partial \log p}{\partial \theta_i} \dfrac{\partial \log p}{\partial \theta_j} \right\rangle_{\text{lin}} \end{array}}$$

$p_2 \to p_1 \equiv p$

Sharma-Mittal $\qquad\qquad\qquad\qquad\qquad$ Supra-extensive

$\log_r \left\langle \dfrac{p_2}{p_1} \right\rangle_{\log_q}$ $\qquad\qquad\qquad\qquad$ $\log_q e_r^{\log \left\langle \frac{p_2}{p_1} \right\rangle_{\log_q}}$

$r \to q \qquad\qquad r \to q$

$r \to 1 \qquad\qquad\qquad\qquad\qquad r \to 1$

Rényi $\qquad\qquad\qquad\qquad\qquad\qquad$ Tsallis

$\log \left\langle \dfrac{p_2}{p_1} \right\rangle_{\log_q}$ $\qquad\qquad\qquad$ $\log_q \left\langle \dfrac{p_2}{p_1} \right\rangle_{\log_q}$

$q \to 1 \qquad\qquad q \to 1$

Kullback-Leibler

$\log \left\langle \dfrac{p_2}{p_1} \right\rangle_{\log}$

$p_2 \to 1$

Shannon (Boltzmann-Gibbs)

$\log \left\langle \dfrac{1}{p_1} \right\rangle_{\log}$

References

[1] S.Amari and H.Nagaoka, *Methods of Information Geometry*, AMS Translations of Mathematical Monographs, vol 191, American Mathematical Society (AMS) and Oxford University Press, 2000.

[2] R. Carrol, *Gravity and the quantum potential*, arXiv:gr-qc/0406004 and *Fluctuations, gravity, and the quantum potential*, arXiv:gr-qc/0501045.

[3] J. M. Corcuera and F. Giummol'e, *A Characterization of Monotone and Regular Divergences*, Ann. Inst. Statist. Math., 50 pp.433-450, 1998.

[4] Fisher, R. A., Philosophical Transaction of the Royal Society of London, (A), Vol.222 (1921), pag. 309-368. See also his further generalizations: Proceedings of the Cambridge Philosophical Society, Vol. 22 (1925), pag. 705-725 and R. A., Proceedings of the Royal Society, (A), vol. 144 (1934), pag. 285-307.

[5] B.R. Frieden, *Physics from Fisher Information: a Unification* (Cambridge Univ. Press, 1998); 2nd ed. *Science from Fisher Information* (Cambridge Univ. Press, June, 2004).

[6] F. Pennini, A. Plastino, *Escort Husimi distributions, Fisher information and nonextensivity*, Phys. Lett. **A**, Vol. 326, 20-26 (2004).

[7] M. Portesi, F. Pennini, A. Plastino, *Geometrical aspects of a generalized statistical mechanics*, cond-mat/0511490.

[8] S. Abe, *Geometry of escort distributions*, Phys. Rev. **E**, 68, 031101 (2003).

[9] J. Naudts, *Estimators, escort probabilities, and phi-exponential families in statistical physics*, J. Ineq. Pure Appl. Math. 5(4), 102 (2004).

[10] J. Naudts, *Continuity of a Class of Entropies and Relative Entropies*, Rev. Math. Phys. 16(6), 809-822 (2004).

[11] P. Jizba, *Information Theory and Generalized Statistics*, in Decoherence and Entropy in Complex Systems: Selected Lectures from DICE 2002. Ed. Hans-Thomas Elze, Lecture Notes in Physics, vol. 633, p.362-376 (2003).

[12] T.D. Frank, A. Daffertshofer, *Exact time-dependant solutions of the Rényi Fokker-Planck equation and the Fokker-Planck equations related to the entropies proposed by Sharma and Mittal*, Physica A **285**, 351 (2000).

[13] T.D. Frank and A.R. Plastino, *Generalized thermostatistics based on the Sharma-Mittal entropy and escort mean values*, Eur. Phys. J. B 30, 543-549 (2002).

[14] B. Lesche, J. Stat. Phys, **27**, 419 (1982).

[15] S. Kullback and R. A. Leibler. *On information and sufficiency.* Ann. Math. Stat. 22(1):7986, 1951.

[16] S. Kullback, *Information Theory and Statistics*, Dover Pubblications, Inc. Mineola, New York.

[17] A.N. Kolmogorov *Sur la notion de la moyenne*, Atti Accad. Naz. Lincei Mem. Cl. Sci. Fis. Mat. Natur. Sez. (6) 12 (1930) 388-391.

[18] M. Masi, *A step beyond Tsallis and Rényi entropies*, Phys. Lett. A, Vol. **338**, 217-224 (2005).

[19] M. Nagumo, *Über eine Klasse der Mittelwerte*, Japan. J. Math., 7 (1930) 7179.

[20] J. Naudts, M. Czachor *Generalized thermostatistics and Kolmogorov-Nagumo averages*, arXiv:cond-mat/0110077 (3 Oct 2001).

[21] Rao, C. R., *Information and accuracy attainable in the estimation of statistical parameters*, Bull. Calcutta Math. Soc. 37, 81-91 (1945), (also published in *Breakthroughs in Statistics*, volume 1: *Foundations and Basic Theory*, New York, Springer-Verlag, pp. 235-48).

[22] A. Rényi, *Probability theory* (North Holland, Amsterdam, 1970); Selected Papers of Alfred Rényi, Vol.2 (Akademia Kiado, Budapest, 1976).

[23] C. C. Rodriguez, *The Metrics Induced by the Kullback Number*, J. Skilling (ed.), Maximum Entropy and Bayesian Methods, pages 415-422, 1989.

[24] Shannon, C. E. e Weaver W., *The Mathematical Theory of Information*, University of Illinois Press, Urbana, Illinois, 1949.

[25] B.D. Sharma and D.P. Mittal, J. Math. Sci. **10**, 28 (1975). See also *New Nonadditive Measures of Relative Information*, J. Comb. Inform. and Syst. Sci., 2, 122-133 (1977).

[26] C. Tsallis, *Possible Generalization of Boltzmann-Gibbs Statistics*, J. Stat. Phys. 52 (1988), 479.

[27] For a general bibliography about nonextensive termodynamics and updates see: http://www.cbpf.br/GrupPesq/StatisticalPhys/TEMUCO.pdf

[28] D. Wagenaar, *Information geometry for neural networks*. Term paper for reading course with A. C. C. Coolen, King's College London, 1998. http://www.its.caltech.edu~pinelab/wagenaar/infogeom.pdf